

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-063298

(43)Date of publication of application : 08.03.1996

(51)Int.Cl.

G06F 3/06

G06F 3/06

(21)Application number : 06-194232

(71)Applicant : HITACHI LTD

(22)Date of filing : 18.08.1994

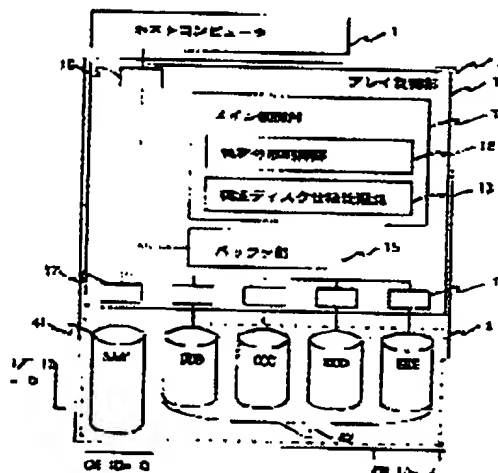
(72)Inventor : HONDA KIYOSHI
ISONO SOICHI
ICHIKAWA MASATOSHI
MATSUMOTO JUN
IWASAKI HIDEHIKO

(54) DISK ARRAY DEVICE

(57)Abstract:

PURPOSE: To improve the system performance of the disk array device consisting of plural disk devices having different specifications by optionally setting the ratio of the number of data stripes according to the individual disk device specifications.

CONSTITUTION: The disk array device 2 consists of an array control part 10 and a disk device group 4 having plural disk devices 41 and 42 having different specifications; and the array control part 10 consists of a main control part 11 which performs total control, a host interface control part 16 which controls an interface with a host 1, plural disk interface control parts 17 which control interfaces with the disk device groups 4, and a buffer part 15 which temporarily stores data transferred between the host 1 and disk devices 41 and 42, and the main control part 11 has at least a load decentralization control part 12 and a constituent disk specification management part 13.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

BEST AVAILABLE COPY

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-63298

(43) 公開日 平成8年(1996)3月8日

(51) Int. Cl.⁹
G 0 6 F 3/06

識別記号 庁内整理番号
5 4 0
3 0 2 A

F I

技術表示箇所

審査請求 未請求 請求項の数21 O L (全 16 頁)

(21) 出願番号 特願平6-194232
(22) 出願日 平成6年(1994)8月18日

(71) 出願人 000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目6番地
(72) 発明者 本田 聖志
神奈川県川崎市麻生区王禅寺1099番地 株
式会社日立製作所システム開発研究所内
(72) 発明者 磯野 聡一
神奈川県川崎市麻生区王禅寺1099番地 株
式会社日立製作所システム開発研究所内
(72) 発明者 市川 正敏
神奈川県川崎市麻生区王禅寺1099番地 株
式会社日立製作所システム開発研究所内
(74) 代理人 弁理士 武 顯次郎

最終頁に続く

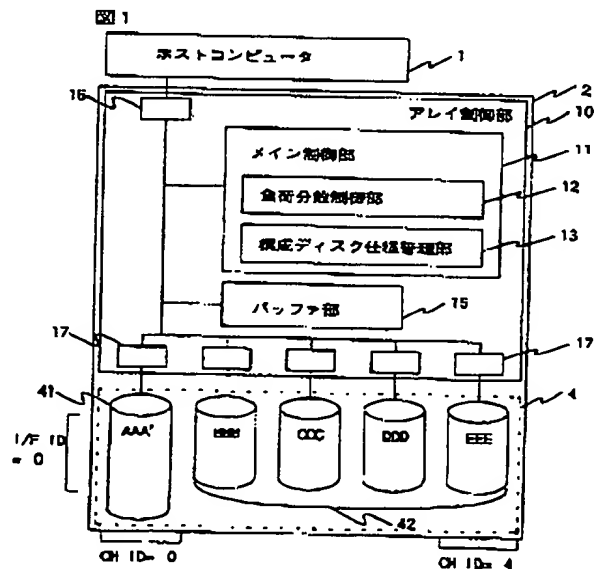
(54) 【発明の名称】 ディスクアレイ装置

(57) 【要約】

【構成】

【目的】 仕様の異なる複数のディスク装置から構成されるディスクアレイ装置において、個々のディスク装置仕様に応じ、データストライプ数比を任意に設定することによって、システム性能の向上を実現すること。

【構成】 ディスクアレイ装置2は、アレイ制御部10と、仕様の異なる複数のディスク装置を有するディスク装置群4から構成され、アレイ制御部は、統括制御を行うメイン制御部11と、ホスト1とのインタフェース制御を行う上位インタフェース制御部16と、ディスク装置群とのインタフェース制御を行う複数のディスクインタフェース制御部17と、ホストとディスク装置間で転送されるデータを一時格納するバッファ部15とから構成され、更に、メイン制御部は、少なくとも負荷分散制御部12と構成ディスク仕様管理部13とを有する。



部に更新・格納する手段とを具備することを特徴とするディスクアレイ装置。

徴とするディスクアレイ装置。

【請求項16】 請求項15記載において、

【特許請求の範囲】

【請求項1】 複数のディスク装置と、該ディスク装置群を制御するアレイ制御部とから構成されるディスクアレイ装置であって、

前記アレイ制御部は、前記上位装置との接続を目的とする上位インタフェース制御部と、前記ディスク装置との接続を目的とするディスクインタフェース制御部と、ディスクアレイ装置の統括制御を行うメイン制御部と、上位装置とディスク装置間での転送データを一時バッファリングするバッファ部とから構成され、

更に、前記メイン制御部において、前記ディスクアレイ装置を構成する個々のディスク装置仕様情報を管理する構成ディスク仕様管理部と、この構成ディスク仕様管理情報に基づき個々のディスク装置に対するアクセス負荷を制御する負荷分散制御部とを具備することを特徴とするディスクアレイ装置。

【請求項2】 請求項1記載において、

前記ディスクアレイ装置を構成する個々のディスク装置仕様に対応して格納ストライプ数を設定することによって、アクセス負荷の分散制御を実現することを特徴とするディスクアレイ装置。

【請求項3】 請求項1記載において、

前記ディスクアレイ装置を構成する個々のディスク装置仕様に対応して格納ストライプ数を設定し、且つ、各々のデータストライプに対する冗長データを格納可能とすることによって、アクセス負荷の分散制御を行う高信頼な記憶装置を実現することを特徴とするディスクアレイ装置。

【請求項4】 請求項3記載において、

前記個々のデータストライプに対する冗長度を確保するため、冗長度の異なる領域を具備することを特徴とするディスクアレイ装置。

【請求項5】 複数のディスク装置と、該ディスク装置群を制御するアレイ制御部とから構成されるディスクアレイ装置であって、

前記アレイ制御部は、前記上位装置との接続を目的とする上位インタフェース制御部と、前記ディスク装置との接続を目的とするディスクインタフェース制御部と、ディスクアレイ装置の統括制御を行うメイン制御部と、上位装置とディスク装置間での転送データを一時バッファリングするバッファ部とから構成され、

更に、前記メイン制御部において、前記ディスクアレイ装置を構成する個々のディスク装置仕様情報を管理する構成ディスク仕様管理部と、前記ディスクアレイ装置を構成する個々のディスク装置に対する過去のアクセスの履歴情報を管理するアクセス履歴情報管理部とを具備し、構成ディスク仕様管理情報及びアクセス履歴管理情報に基づき個々のディスク装置に対するアクセス負荷を制御する負荷分散制御部とを具備することを特徴とするディスクアレイ装置。

【請求項6】 請求項5記載において、

前記ディスクアレイ装置を構成する個々のディスク装置仕様情報に対応して格納ストライプの設定・再配置を行うことによって、アクセス負荷の分散制御を実現することを特徴とするディスクアレイ装置。

【請求項7】 請求項5記載において、

前記ディスクアレイ装置を構成する個々のディスク装置仕様情報及び過去のアクセス履歴情報に対応して格納ストライプ数を設定し、且つ、各々のデータストライプに対する冗長データを格納可能とすることによって、アクセス負荷の分散制御を行う高信頼な記憶装置を実現することを特徴とするディスクアレイ装置。

【請求項8】 請求項5 または6 または7 記載において、

前記ディスクアレイ装置を構成する個々のディスク装置仕様情報及び過去のアクセス履歴管理情報に基づき、格納ストライプの設定・再配置を行う際、ライトアクセス頻度の高いデータを冗長度の高い領域に再配置することを特徴とするディスクアレイ装置。

【請求項9】 請求項5 または6 または7 記載において、

前記ディスクアレイ装置を構成する個々のディスク装置仕様情報及び過去のアクセス履歴管理情報に基づき、格納ストライプの設定・再配置を行なう際、冗長度の異なる各領域毎に、格納ストライプの設定・再配置アルゴリズムを、ユーザが任意に設定可能とすることを特徴とするディスクアレイ装置。

【請求項10】 複数のディスク装置と、該ディスク装置群を制御するアレイ制御部とから構成され、

前記アレイ制御部は、前記上位装置との接続を目的とする上位インタフェース制御部と、前記ディスク装置との接続を目的とするディスクインタフェース制御部と、ディスクアレイ装置の統括制御を行うメイン制御部と、上位装置とディスク装置間での転送データを一時バッファリングするバッファ部と、格納データから冗長データを生成する冗長データ生成部とから構成され、前記ディスクアレイ装置を構成する個々のディスク装置には、データと各々のデータに対応する冗長データとを格納し、

前記ディスクアレイ装置を構成するディスク装置に障害が発生した場合、前記冗長データを用いて障害ディスク装置のデータを回復可能とするディスクアレイ装置であって、

前記メイン制御部において、前記ディスクアレイ装置を構成する個々のディスク装置仕様情報を管理する構成ディスク仕様管理部と、この構成ディスク仕様管理情報に基づき個々のディスク装置に対するアクセス負荷を制御する負荷分散制御部と、前記障害ディスク装置の交換及び交換されたディスク装置仕様を検出する手段と、この交換ディスク装置仕様情報を前記構成ディスク仕様管理

5

【産業上の利用分野】本発明はディスクアレイ装置に係り、特に、仕様の異なる複数のディスク装置から構成されるディスクアレイ装置のシステム構成及び制御方式に関する。

【0002】

【従来の技術】従来、記憶装置の性能を改善することを目的として、複数のディスク装置を並列のインタフェースを介して接続することによって、アクセス時間の短い大容量のディスクサブシステムを構築するディスクアレイ装置が開発／製品化されている。

【0003】特に、上記ディスクアレイ (Redundant Arrays of Inexpensive Disks) 装置の構成及びその制御方式に関しては、特開平2 - 2 3 6 7 1 4 号公報に、任意台数のディスク装置から構築可能とするシステムが開示されている。

【0004】この先願に開示された従来技術におけるディスクアレイ装置は、同一仕様の複数のディスク装置から構成され、個々のディスク装置の物理的配置には何等制限を持たないというものである。

【0005】また、特開平6 - 7 5 7 0 7 号公報には、容量の異なる複数のディスク装置からディスクアレイ装置を構築可能とするシステムが開示されている。

【0006】

【発明が解決しようとする課題】上記した前者の従来技術によれば、仕様を同一とする複数のディスク装置から構築されるディスクアレイ装置において、高性能なシステムを提供しているが、該ディスクアレイ装置は、全て同一のディスク装置を用いて構築しなければならないという問題が有る。これは、ディスク装置単体のビットコストの低下に伴う大容量・低価格化及び高性能化がめざましい現在、ディスクアレイ装置を構成する1台のディスク装置に障害が発生した場合、数年前の容量の小さい性能の低い同一仕様のディスク装置を用いて、ディスクアレイ装置を再構築することをユーザに強いるということであり、非常に不経済であるといえる。

【0007】また、前記した後者の従来技術は、上記の問題を解決し、ディスクアレイ装置の構成をより柔軟にするものであるが、性能に関する考慮が十分ではなかった。

【0008】つまり、前記後者の従来技術は、データストライプのマッピング処理を個々のディスク装置とは独立に実現することによって、最新の大容量ディスク装置等の容量の異なる複数のディスク装置からディスクアレイ装置を構築可能とするものである。しかし、上記のマッピング処理において、データストライプの最適な配置を実現するものではなく、前記ディスクアレイ装置を構成する個々のディスク装置仕様 (最新の高性能ディスク装置仕様) を十分に活用するものではないといえる。

【0009】すなわち、従来方式のディスクアレイ装置においては、ディスクアレイ装置を構成するディスク装

6

置の1台を、容量及び性能が2倍の他のディスク装置に交換した場合、この容量が2倍のディスク装置のストライプ数を、他のディスク装置に対し2倍にすることは不可能である。なぜなら、冗長データとしてのパリティを格納することによって信頼性を向上するRAID3, 4, 5のシステムにおいては、パリティストライプは各々のディスク装置の1つのデータストライプと対応付けられる必要があるため、各カラムを構成するディスク装置の総ストライプ数は等しくなければならないという制限によるからである。上記理由のため、従来方式では、上記のように容量が2倍のディスク装置においてもその1/2の領域しか使用できず無駄にしてしまうか、この領域を他のディスク装置として使用する場合においては、冗長度のない信頼性の低い領域としてしか使用できないことになる。

【0010】また、各カラムを構成するディスク装置の総ストライプ数が等しいということから、個々のディスク装置 (カラム) に対するアクセス発生頻度も等しくなると考えられる。このため、2倍の性能を持つディスク装置を用いてシステムを構築しても、その効果が期待されないという問題がある。

【0011】以上のことから、前記従来技術を採用した容量及び性能の異なる複数のディスク装置から構築されるディスクアレイ装置は、ディスク装置の容量及び性能差を十分考慮したものではなく、従来のディスクアレイ装置の一部に大容量・高性能なディスク装置を用いても、他のディスク装置のためシステム性能が向上しない、或いは記憶容量が増大しないという問題がある。

【0012】本発明の目的は、上記した従来技術における問題点を解消し、ディスクアレイ装置を構成する個々のディスク装置仕様及び過去のアクセス傾向に応じて、データストライプ或いは該データストライプに対応する冗長データストライプの再配置を行い、個々のディスク装置へのアクセス負荷を最適化することによって、システム性能を向上させることにある。

【0013】

【課題を解決するための手段】上記目的を達成するため、本発明におけるディスクアレイ装置は、複数のディスク装置と、該ディスク装置群を制御するアレイ制御部とから構成され、更に、前記アレイ制御部は、上位装置との接続を目的とする上位インタフェース制御部と、前記ディスク装置との接続を目的とするディスクインタフェース制御部と、ディスクアレイ装置の統括制御を行うメイン制御部と、上位装置とディスク装置間での転送データを一時バッファリングするバッファ部とから構成される。

【0014】また、前記アレイ制御部は、格納データから冗長データを生成する冗長データ生成部を有し、更に、該冗長データ生成部は、障害ディスク装置発生時に、冗長データと他のディスク装置のデータとから、障

害ディスク装置に格納されていたデータを回復する機能を有する。

【0015】更に、前記メイン制御部は、前記ディスクアレイ装置を構成する個々のディスク装置仕様情報を管理する構成ディスク仕様管理部と、この構成ディスク仕様管理情報に基づき個々のディスク装置に対するアクセス負荷を制御する負荷分散制御部とを有する。

【0016】また、前記メイン制御部は、前記ディスクアレイ装置を構成する個々のディスク装置に対する過去のアクセスの履歴情報を管理するアクセス履歴情報管理部と、前記ディスクアレイ装置を構成するディスク装置がダウンした場合、障害ディスク装置の交換及び交換されたディスク装置仕様を検出し、これを前記構成ディスク仕様管理部に更新格納する手段と、前記ディスクアレイ装置を構成するディスク装置を増設した場合、増設されたディスク装置仕様を検出し、これを前記構成ディスク仕様管理部に新規に格納する手段と、を有するようにもされる。

【0017】

【作用】前記上位インタフェース制御部は、アレイ制御部と上位装置とを接続し、前記ディスクインタフェース制御部は、アレイ制御部とディスク装置とを接続する。

【0018】前記アレイ制御部は、ディスクアレイ装置の統括制御を行うメイン制御部と、上位装置とディスク装置間での転送データを一時バッファリングするバッファ部とを具備し、上記メイン制御部における統括制御では、マッピング処理、キャッシュ管理等が行われる。

【0019】また、アレイ制御部の前記冗長データ生成部においては、格納データから冗長データ（パリティあるいはミラーデータ）の生成処理、及び障害ディスク装置発生時に、障害ディスク装置に格納されていたデータの回復処理を実行する。

【0020】更に、メイン制御部における前記構成ディスク仕様管理部は、ディスクアレイ装置を構成する個々のディスク装置仕様情報を格納/管理し、前記負荷分散制御部は、上記の構成ディスク仕様管理情報や後述のアクセス履歴管理情報等に基づき、個々のディスク装置に対するアクセス負荷を制御する。

【0021】また、メイン制御部における前記アクセス履歴情報管理部は、ディスクアレイ装置を構成する個々のディスク装置に対する過去のアクセス履歴管理情報を格納/管理する。

【0022】更に、前記メイン制御部は、前記ディスクアレイ装置を構成するディスク装置がダウンした場合、障害ディスク装置の交換及び交換されたディスク装置仕様を検出し、これを前記構成ディスク仕様管理部に更新格納し、また、前記ディスクアレイ装置を構成するディスク装置を増設した場合、増設されたディスク装置仕様を検出し、これを前記構成ディスク仕様管理部に新規に格納する。

【0023】

【実施例】以下、本発明を図示した各実施例によって説明する。

【0024】〈第1実施例〉図1は、本発明の第1実施例に係るディスクアレイ装置のブロック図である。同図において、1は上位装置たるホストコンピュータ（以下、ホスト1と称す）、2はディスクアレイ装置、4は複数のディスク装置4.1及び4.2…で構成されるディスク装置群、10はアレイ制御部であり、ディスクアレイ装置2は、ディスク装置群4とアレイ制御部10とで構成される。11は、ディスクアレイ装置2の統括制御を行うアレイ制御部10のメイン制御部で、負荷分散制御部12や構成ディスク仕様管理部13などを具備している。15はアレイ制御部10のバッファ部で、ホスト1とディスク装置群4との間での転送データを一時バッファリングする。16はアレイ制御部10の上位インタフェース制御部で、ホスト1とアレイ制御部10との間のインタフェースを司る。17…はアレイ制御部10のディスクインタフェース制御部で、アレイ制御部10と各ディスク装置4.1及び4.2…との間のインタフェースを司る。

【0025】上記したアレイ制御部10は、図示していないが、ディスク装置群4の格納データから冗長データ（RAID3、4、5ではパリティ、RAID1ではミラーデータ）を生成して所定のディスク装置に格納させる冗長データ生成部を有し、この冗長データ生成部は、障害ディスク装置発生時に、冗長データと他のディスク装置のデータとから、障害ディスク装置に格納されていたデータを回復する機能も具備している。また、メイン制御部11による統括制御では、マッピング処理、キャッシュ管理等が行われる。

【0026】本実施例は、アレイ制御部10のメイン制御部11に、負荷分散制御部12と構成ディスク仕様管理部13とを具備し、仕様（性能、容量）の異なる複数のディスク装置4.1及び4.2…から構成されるディスクアレイ装置2において、ディスクアレイ装置2を構成する個々のディスク装置（4.1、4.2）仕様に基づき、個々のディスク装置に対するストライプ数の設定による負荷分散制御、及びデータ、冗長データの配置の最適化を行うことによって、システム性能の向上を実現するものである。

【0027】以下、何らかの要因（例えば、ディスク装置の障害発生等）で、ディスクアレイ装置2を構成する1台のディスク装置が、他のディスク装置4.2と仕様の異なるディスク装置4.1に交換され、このディスク装置4.1が他のディスク装置4.2に対し、ディスク容量及び性能が2倍程度である場合を例にとって、本実施例を説明する。

【0028】ここで、前記ディスク装置群4は、インタフェースID（以下、I/F IDと称す）=0、チャ

9

ンネルID(以下、CH IDと称す) = 0 のディスク装置41 (ID=AAA') と、該ディスク装置41と仕様の異なる他のディスク装置42 (ID=BBB~EEE) とから構成されており、上記したように、ディスク装置41の装置仕様が他のディスク装置42の装置仕様に対し、ディスク容量及び性能が2倍程度 ($x \times 1 = x \times 0 \times 2$, $s \times 1 = s \times 0 / 2$, $r \times 1 = r \times 0 \times 2$, $t \times 1 = t \times 0 \times 2$; 図2参照) となっている。

【0029】図2は、前記構成ディスク仕様管理部13における個々の構成ディスク仕様管理情報の1例を示す図である。管理情報としては、I/F ID31と、CH ID32と、ディスクID33と、ディスク容量34と、平均シーク時間35と、回転数36と、データ転送性能37とを有している。但し、構成ディスク仕様管理情報はこの例に限定されるものではなく、これらの一部の情報を用いて本発明を実施することには何等問題ないことはいふまでもない。

【0030】図3は、図1に示したディスク装置群4における、データ及び該データに対応する冗長データの配置の1例を示す図である。図3に示した例では、ディスク装置群4は、冗長度の異なる2つの領域(RAID5の領域20と擬似RAID1の領域21)から構成されており、データ22に対しパリティ23を格納することによって、冗長度を確保し、信頼性の高い記憶装置を実現するようになっている。なお、図3の各ディスク装置(41, 42-1, 42-2, 42-3, 42-4)内に記載された番号は、ストライプ番号を示している。

【0031】ここで、RAID5の領域においては、ストライプ番号「0」～「3」のデータに対するパリティは、ストライプ番号「P0」である。また、擬似RAID1の領域における冗長データは、パリティとミラーデータであり、例えば、ストライプ番号「16」のデータに対する冗長データは、ストライプ番号「16'」(ミラーデータ)及びストライプ番号「P4」(パリティ)である。

【0032】本実施例においては、ストライプ番号「16」～「19」を正データ(通常アクセスされるデータ)、ストライプ番号「16'」～「19'」を副データ(通常アクセスされないデータ)とすることによって、ディスク装置41へのアクセス発生頻度を、他のディスク装置42に対して増加させるようにしてある。

【0033】なお、本発明の各実施例で用いる上記擬似RAID1は、通常のRAID1レベルのように、冗長度を確保するためにミラーデータのみをもつもの(ディスクデータを単に2重化したもの)とは異なり、上述したように、冗長データとしてミラーデータとパリティとをもち、見掛け上は通常のRAID1のようにふるまうことが可能であるも、必要に応じ適宜パリティを使用することも可能となっており、したがって、通常のRAID1レベルよりも高信頼なRAIDとして構築されてい

10

る。勿論、本発明で通常のRAID1レベルを用いても何等差支えない。

【0034】図4は、従来の全て同一仕様のディスク装置から構成されたディスクアレイ装置と、その1台を高性能ディスク装置に交換した本実施例のディスクアレイ装置との、ランダムリードアクセス時の性能比較を示す1例であり、図中の波線はシーク/回転待ち時間を、実線はデータ転送時間をそれぞれ示している。

【0035】本実施例において、I/F ID=0, CH ID=0のディスク装置をI D=AAA'のディスク装置41に交換した際には、前記アレイ制御部10のメイン制御部11は、次の①～④の処理を実行する。

【0036】①. 交換ディスク装置の検出
アレイ制御部10のメイン制御部11において、I/F ID=0, CH ID=0のディスク装置の交換を検出する。

【0037】②. 交換ディスク装置仕様の獲得
メイン制御部11の構成ディスク仕様管理部13は、交換したI/F ID=0, CH ID=0のディスク装置仕様を獲得し、構成ディスク仕様管理情報を更新・管理する。

【0038】③. 負荷分散制御方式の決定
メイン制御部11の負荷分散制御部12は、上記した構成ディスク仕様管理部13によって管理される情報に基づき、ストライプ数の設定を行う。本実施例では、ストライプ数の設定によって負荷分散制御を行うものであり、交換されたディスク装置41の仕様が、他のディスク装置42の仕様の2倍であるとしていることから、交換されたディスク装置41に対し2倍のストライプ数を設定する。

【0039】しかし、従来方式のディスクアレイ装置においては、前述したように、ディスク装置41のストライプ数を、他のディスク装置42に対し2倍にすることは不可能であった。そのため、従来方式では、本実施例の構成におけるストライプ番号「16」～「19」の領域を無駄にしてしまうか、冗長度のない信頼性の低い領域としてしか使用できなかった。また、ディスク装置41, 42の総ストライプ数が等しいということから、個々のディスク装置に対するアクセス発生頻度も等しくなり、2倍の性能を持つディスク装置41を用いてシステムを構築しても、その効果が期待されないという問題があった。

【0040】そこで、本発明では、複数のRAIDレベルからディスクアレイ装置2を構築可能とすることによって、上記の問題を解決し、仕様の異なる複数のディスク装置から構成される前記ディスクアレイ装置2の容量及び性能スケラビリティを実現するものである。本実施例では、上記複数のRAIDレベルとして、擬似RAID1とRAID5を採用しているが、これに限定されるものではないことはいふまでもない。

11

【 0 0 4 1 】 ④. データの回復・格納処理

メイン制御部11において、前記構成ディスク仕様管理部13及び前記負荷分散制御部12からの情報に基づき、交換されたディスク装置41に対するデータの回復処理を、以下の手順で実行する。

【 0 0 4 2 】 (1) RAI D5の再構築

データストライプ(#0 ~ #19) 及び各々に対応するパリティストライプ(#P0 ~ #P4) に関し、従来通りのデータ回復処理によって、ディスク装置41のデータ回復を実現する。すなわち、各ディスク装置42 (I D = B B B ~ E E E) のデータ及びパリティから、ディスク装置41 (I D = A A A ') のデータストライプ及びパリティストライプを回復・格納することによって、RAI D5システムの再構築を実現する。

【 0 0 4 3 】 (2) 擬似RAI D1の構築

ディスク装置41に格納されたパリティストライプ(#P4) に対応するデータストライプ(#16 ~ #19) をディスク装置41に格納し、各ディスク装置42のデータストライプ(#16' ~ #19') をミラーデータとすることによって、擬似RAI D1システムを構築する。

【 0 0 4 4 】 以上の処理を実行することによって、RAI Dレベルの異なる複数の領域(RAI D5と擬似RAI D1) から構成され、ディスク装置41の総データストライプ数をディスク装置42の2倍とし、総データストライプ数比と性能比とを同等とするディスクアレイ装置2が構築可能となる。

【 0 0 4 5 】 このことから、ディスク装置41へは、ディスク装置42に較べ2倍のアクセス発生が期待される。すなわち、従来方式では、全てのディスク装置に対して等しい確率でアクセス要求が発行される場合、ディスク装置41の性能が活ききれず任意の処理時間内に5件のI/O要求しか処理されないのに対し、本実施例のシステムによれば、ディスク装置41へのアクセス発生が他のディスク装置に対するアクセス発生2倍になることが期待され、上記任意の処理時間内に6件のI/O要求が処理可能となって、1.2倍のシステム性能の向上が図られる。

【 0 0 4 6 】 図5は、図3の状態から、更にI/F I D=0, CH I D=1のディスク装置が交換され、ディスクアレイ装置2を構成する複数のディスク装置のうち、2台のディスク装置の仕様が他のディスク装置の仕様と異なる場合の、データ及び該データに対応する冗長データの配置の1例を示す図である。図5において、ディスク装置群4は、図3と同様に、冗長度の異なる2つの領域(RAI D5の領域20と擬似RAI D1の領域21) から構成されている。

【 0 0 4 7 】 I/F I D=0, CH I D=1のディスク装置が交換された場合にも、メイン制御部11は、先と同様に、以下の①~④の処理を実行する。

12

【 0 0 4 8 】 ①. 交換ディスク装置の検出

メイン制御部11において、I/F I D=0, CH I D=1のディスク装置の交換を検出する。

【 0 0 4 9 】 ②. 交換ディスク装置仕様の獲得

メイン制御部11の構成ディスク仕様管理部13は、交換したI/F I D=0, CH I D=1のディスク装置仕様を獲得し、構成ディスク仕様管理情報を更新・管理する。

【 0 0 5 0 】 ③. 負荷分散制御方式の決定

メイン制御部11の負荷分散制御部12は、上記した構成ディスク仕様管理部13によって管理される情報に基づき、ストライプ数の設定を行う。ここでも交換されたディスク装置41(41-1) の仕様が、元からある他のディスク装置42の仕様の2倍であるものとし、交換されたディスク装置41-1に対し2倍のストライプ数を設定する。

【 0 0 5 1 】 ④. データの回復・格納処理

メイン制御部11において、前記構成ディスク仕様管理部13及び前記負荷分散制御部12からの情報に基づき、交換されたディスク装置41-1に対するデータの回復処理を、以下の手順で実行する。

【 0 0 5 2 】 (1) RAI D5の再構築

データストライプ(#0 ~ #19) 及び各々に対応するパリティストライプ(#P0 ~ #P4) に関し、従来通りのデータ回復処理により、ディスク装置41-1 (I D = B B B ') のデータ回復を実現する。すなわち、ディスク装置41(41-0 ; I D = A A A ') と各ディスク装置42 (I D = C C C ~ E E E) のデータ及びパリティから、ディスク装置41(41-1 ; I D = B B B ') のデータストライプ及びパリティストライプを回復・格納することによって、RAI D5システムの再構築を実現する。

【 0 0 5 3 】 (2) 擬似RAI D1の構築

ディスク装置41(41-1 ; I D = B B B ') に格納されたパリティストライプ(#P3) に対応するデータストライプ(#12 ~ #15) を、ディスク装置41(41-1 ; I D = B B B ') に格納し、ディスク装置41(41-0 ; I D = A A A ') , 各ディスク装置42 (I D = C C C ~ E E E) のデータストライプ(#12' ~ #15') を、ミラーデータとすることによって、擬似RAI D1システムを構築する。

【 0 0 5 4 】 以上の処理を実行することによって、図5に示すRAI Dレベルの異なる複数の領域から構成されるデータ及び該データに対応する冗長データの配置を実現可能とし、ディスク装置41の総データストライプ数(=7) を、ディスク装置42の総データストライプ数(=2) の3.5倍とすることから、ディスク装置仕様比(最大3.5倍) までシステム性能比が向上すること、並びに、同一データを複数のディスク装置に格納することから、多重度の向上によるシステム性能の向上が

期待されるようになる。

【0055】図6は、図5と同様に、ディスクアレイ装置2を構成する複数のディスク装置のうち、2台のディスク装置の仕様が他のディスク装置の仕様と異なる場合の、データ及び該データに対応する冗長データの配置の他の1例を示す図である。図6において、ディスク装置群4は、図3及び図5の場合と同様に、冗長度の異なる2つの領域(RAID5の領域20と擬似RAID1の領域21)から構成されているが、本例では、2台のディスク装置41(41-0, 41-1)のみで擬似RAID1の領域21を構成している。また、擬似RAID1の領域21において、ミラーデータ(#20' ~ #23')は各々のディスク装置41-0, 41-1に均等に格納配置するようにしている。

【0056】本例を実現する場合、メイン制御部11は、先に述べたのと同様な、①. 交換ディスク装置の検出、②. 交換ディスク装置仕様の獲得、③. 負荷分散制御方式の決定、④. データの回復・格納処理という一連の処理を実行し、④のデータの回復・格納処理では、先の例と同様に、まず、RAID5の領域のデータの再構築を行い、次に、2台のディスク装置41-0(ID=AAA'), 41-1(ID=BBB')でRAID5に使用されていない領域21を、擬似RAID1の領域として新たに構築する。

【0057】以上の処理を実行することによって、図6に示すRAIDレベルの異なる複数の領域から構成されるデータ及び該データに対応する冗長データの配置を実現可能とし、ディスク装置41の総データストライプ数(=6)をディスク装置42の総データストライプ数(=4)の1.5倍とすることから、ディスク装置仕様比(最大1.5倍)までシステム性能比が向上すること、並びに、同一データを複数のディスク装置に格納することから、多重度の向上によるシステム性能の向上が期待されるようになる。

【0058】更にまた、図6に示した本例によれば、ディスクアレイ装置2の総記憶容量は、従来方式及び図3、図5の例に対し1.2倍とすることが可能であり、容量の大きなディスク装置を使用した効果を活かすことが可能となる。なお、本例では、ミラーデータ(#20' ~ #23')は各々のディスク装置41-0, 41-1に均等に格納配置するものとしているが、これに限るものではない。

【0059】図7は、図6の状態から更にI/F ID=0, CH ID=2のディスク装置が交換され、ディスクアレイ装置2を構成する複数のディスク装置のうち、3台のディスク装置の仕様が他のディスク装置の仕様と異なる場合の、データ及びデータに対応する冗長データの配置の1例を示す図である。図7において、ディスク装置群4は、冗長度の異なる2つのRAID5の領域24, 25から構成されている。

【0060】本例を実現する場合、メイン制御部11は、先に述べたのと同様な、①. 交換ディスク装置の検出、②. 交換ディスク装置仕様の獲得、③. 負荷分散制御方式の決定、④. データの回復・格納処理という一連の処理を実行する。そして、④のデータの回復・格納処理では、まず、RAID5の領域24のデータの再構築を行い、次に、各ディスク装置41-0(ID=AAA'), 41-1(ID=BBB'), 41-2(ID=CCC')で前記領域24に使用されていない領域25を、RAID5の領域として新たに構築する。以下に、ディスク装置41-0(ID=AAA'), 41-1(ID=BBB'), 41-2(ID=CCC')で前記領域24に使用されていない領域25を、RAID5の領域として新たに構築する処理手順の1例を示す。

【0061】(1) RAID5の再構築
ディスク装置41-0(ID=AAA'), 41-1(ID=BBB')とディスク装置42-3(ID=DDD), 42-4(ID=EEE)のデータ及びパリティから、新たに交換されたディスク装置41-2(ID=CCC')のデータストライプ(#2, #6, #13, #17)及びパリティストライプ(#P2)を回復・格納することによって、前記領域24のRAID5の再構築を行なう。

【0062】(2) 新規RAID5領域の構築
データストライプ(#24 ~ #27)を、新たに交換されたディスク装置41-2(ID=CCC')に格納し、図6におけるミラーデータ(#20' ~ #23')をパリティに置き換える(例えば、#20, #24から#Pn0を生成する)ことによって、前記領域25のRAID5システムを新たに構築する。

【0063】以上の処理を実行することによって、図7に示したRAIDレベルの異なる複数の領域から構成されるデータ及び該データに対応する冗長データの配置を実現可能とし、ディスク装置41の総データストライプ数(=6)をディスク装置42の総データストライプ数(=4)の1.5倍とすることから、システム性能比が向上すること、並びに、同一データを複数のディスク装置に格納することから、多重度の向上によるシステム性能の向上が期待されるようになる。

【0064】更に、図7に示した本例によれば、ディスクアレイ装置2の総記憶容量は、従来方式、及び前述したRAID5の領域に使用されていない領域をRAID1の領域21として使用する例に対し、冗長度を下げることによって1.6倍とすることが可能であり、容量の大きなディスク装置を使用した効果を充分活かすことが可能となる。なお、本例では、ミラーデータ(#20' ~ #23')をパリティデータ(#Pn0 ~ #Pn3)に置き換えるものとしているが、これに限るものではない。更にまた、本例における前記領域24は、前述の各例と同様に擬似RAID1(もしくは通常のRAID

15

1) の領域として、あるいは他のRAIDレベル(3または4)の領域として使用することも可能である。

【0065】以上の各例において、ディスクアレイ装置2を、4台のデータディスク装置と1台のパリティディスク装置から構成するものとしているが、これに限定されるものではない。更に、例えば前記した図3の構成において、交換後のディスク装置41にパリティストライプ(#P4)を回復・格納しているため、ディスク装置41は少なくともディスク装置42の1.8倍のディスク容量が要求されるが、このパリティストライプ(#P4)は、特に回復・格納する必要はない。こうした場合には、ディスク装置41に対するディスク容量の制限を低くすること、すなわち、ディスク容量の少ないディスク装置41(ディスク装置42の1.6倍のディスク容量のディスク装置41)を用いることが可能となる。

【0066】また、本実施例の各例では、交換後のディスク装置41の性能比を元からあるディスク装置42の2倍とした場合を例にとったが、これに限定されるものではなく、任意の性能比を境界値として使用することが可能である。

【0067】更にまた、データストライプの配置は、本実施例の各例に限られるものではなく、後述の通り、他の配置方式の採用及び再配置を行うことも可能である。

【0068】また、本実施例において述べてきた事項(ディスクアレイ構成、性能仕様比、ストライプの配置方式等)は、後述の実施例においてもなんら制限されるものではない。

【0069】(第2実施例)図8は、本発明の第2実施例に係るディスクアレイ装置のブロック図であり、同図において、図1の第1実施例と均等な構成要素には同一符号を付し、その説明は重複を避けるため割愛する。図8において、14は前記メイン制御部11に設けられたアクセス履歴情報管理部である。

【0070】本実施例は、アレイ制御部10のメイン制御部11に、負荷分散制御部12と構成ディスク仕様管理部13とアクセス履歴管理部14とを具備し、仕様(性能、容量)の異なる複数のディスク装置41、42から構成されるディスクアレイ装置2において、該ディスクアレイ装置2を構成する個々のディスク装置(41、42)仕様、及び個々のディスク装置(41、42)に対する過去のアクセス履歴情報に基づき、データ、冗長データの配置の最適化を行うことによって、システム性能の向上を実現するものである。

【0071】図9は、上記アクセス履歴管理部14における、個々の構成ディスクに対する過去のアクセス履歴情報の1例を示す図である。本実施例においては、データストライプ毎に、リードアクセス回数(RD0~RD23)、ライトアクセス回数(WR0~WR23)、トータルのアクセス回数(TT0~TT23)、及び各々についてのアクセス発生頻度順位で管理するよ

16

うにしている。但し、アクセス履歴管理情報は、この例に限定されるものではなく、これらの一部の情報或いは同類の情報(ファイル、パーティション、ディスク装置毎のアクセス履歴管理情報)を用いて本発明を実施することは何等問題ないことは明かである。

【0072】図10は、ディスクアレイ装置2を構成する複数のディスク装置のうち、1台のディスク装置の仕様が他のディスク装置の仕様と異なる場合の、データ及び該データに対応する冗長データの配置の1例を示す図である。図10において、ディスク装置群4は、前記図3の例と同様に、冗長度の異なる2つの領域(RAID5の領域20と擬似RAID1の領域21)から構成されている。

【0073】以下、第1実施例と同様に、何らかの要因でディスクアレイ装置2を構成するI/F ID=0, CH ID=0のディスク装置(ID=AAA)が、他のディスク装置42(ID=BBB~EEE)と仕様の異なるディスク装置41(ID=AAA')に交換され、該ディスク装置41が他のディスク装置42に対し、ディスク容量及び性能が2倍である場合($x \times 1 = x \times 0 \times 2$, $ss1 = ss0 / 2$, $rr1 = rr0 \times 2$, $tt1 = tt0 \times 2$; 図2参照)を例にとって、本実施例を説明する。

【0074】本実施例においても、I/F ID=0, CH ID=0のディスク装置がID=AAA'のディスク装置41に交換された際には、アレイ制御部10のメイン制御部11は、次の①~④の処理を実行する。

【0075】①. 交換ディスク装置の検出
アレイ制御部10のメイン制御部11において、I/F ID=0, CH ID=0のディスク装置の交換を検出する。

【0076】②. 交換ディスク装置仕様の獲得
メイン制御部11の構成ディスク仕様管理部13は、交換したI/F ID=0, CH ID=0のディスク装置仕様を獲得し、前記した構成ディスク仕様管理情報を更新・管理する。

【0077】③. 負荷分散制御方式の決定
負荷分散制御部12は、上記した構成ディスク仕様管理部13によって管理される情報に基づき、ストライプ数の設定を行う。本実施例においても、ストライプ数の設定によって負荷分散制御を行うものであり、交換されたディスク装置41の仕様が、他のディスク装置42の仕様の2倍であるとしていることから、交換されたディスク装置41に対し2倍のストライプ数を設定する。

【0078】なお本実施例でも、前述の第1実施例と同様に、複数のRAIDレベルとしてRAID5及び擬似RAID1を採用しているが、これに限定されるものではないことはいうまでもない。

【0079】④. データの回復・格納処理
メイン制御部11において、前記構成ディスク仕様管理

部13及び前記負荷分散制御部12からの情報に基づき、交換されたディスク装置に対するデータの回復処理を、以下の①～④の手順で実行する。

【0080】(1) RAI D5の再構築

データストライプ(#0～#19)及び各々に対応するパリティストライプ(#P0～#P4)に関し、従来通りのデータ回復処理によって、ディスク装置41のデータ回復を実現する。すなわち、各ディスク装置42(ID=BBB～EEE)のデータ及びパリティから、ディスク装置41(ID=AAA')のデータストライプ及びパリティストライプを回復・格納することによって、RAI D5システムの再構築を実現する。

【0081】(2) 擬似RAI D1の構築

ディスク装置41に格納されたパリティストライプ(#P4)に対応するデータストライプ(#16～#19)をディスク装置41に格納し、各ディスク装置42のデータストライプ(#16'～#19')をミラーデータとすることによって、擬似RAI D1システムを構築する。

【0082】以上の処理を実行することによって、前記した図3に示すデータ及び該データに対応する冗長データの配置が実現される。本実施例では、この後さらに、個々のデータストライプに対する過去のアクセス履歴管理情報に基づき、データストライプの最適配置を実現するものであり、以下に、このデータストライプの再配置処理を示す。

【0083】(3) データストライプの再配置

図9に示したアクセス履歴管理情報例によれば、データストライプ(#14～#17)へのライトアクセス頻度が高いことがわかる。RAI D5においては、パリティストライプのデータを生成する処理のため、リード処理に対しライト処理の性能が低下することが知られている。そこで、上記ライトアクセス頻度の高いデータストライプ(#14～#17)を、ライト処理時パリティストライプデータを生成する必要のない擬似RAI D1の領域のデータストライプ(#16～#19)が格納されていた位置に格納することが望ましい。ここで、データストライプ(#16、#17)は、すでに擬似RAI D1領域に格納されていることから、データストライプ(#14、#15)を擬似RAI D1領域に移動すれば良いことが解る。

【0084】このため、以下のa～cの処理が実行される。

【0085】a. 図3のデータストライプ(#14、#15)を、図3のデータストライプ(#18、#19)が格納されていた位置にコピーする。

b. 図3のデータストライプ(#18'、#19')を、図3のデータストライプ(#14、#15)が格納されていた位置に、データストライプ(#18、#19)として移動する。

c. データストライプ(#14、#15)を、図3のデータストライプ(#18'、#19')が格納されていた位置に、データストライプ(#14'、#15')としてコピーする。

【0086】以上の処理手順によって、データストライプ(#14、#15、#18、#19)の図10に示したような再配置が実現される。

【0087】次に、図9のアクセス履歴管理情報例では、データストライプ(#4)のリードアクセス頻度が高いことから、このデータストライプ(#4)を高性能なディスク装置41に格納することが望ましく、上記データストライプ(#4)とデータストライプ(#7)とが置き換えられる。

【0088】(4) パリティストライプの再構築

データストライプ(#14、#15、#18、#19)を再配置したことによって、対応するパリティストライプ(#P3、#P4)の再構築が必要である。これは、対応するデータストライプ(#12、#13、#18、#19、及び#14'～#17')から、パリティストライプ(#P3'、#P4')として生成され、格納される。

【0089】以上の処理を実施することによって、仕様(性能、容量)の異なる複数のディスク装置から構成されるディスクアレイ装置において、該ディスクアレイ装置を構成する個々のディスク装置仕様、及び個々のディスク装置に対する過去のアクセス履歴情報に基づいた、データ、冗長データの配置の最適化を行うことが可能となり、システム性能の向上を実現できる。

【0090】なお、本実施例においては、装置仕様の異なるディスク装置に交換後、データの再配置を実施しているが、当然再配置以前にデータの最適配置を行うことが望ましい。このとき、前記ディスクアレイ装置を構成するディスク装置が全て同一の場合は、個々のディスク装置に対するトータルのアクセスを分散させるように再配置することが望ましい。

【0091】また、本実施例では、ライトのアクセス頻度を優先したデータストライプの配置を行っているが、これに限るものではない。例えば、リード処理性能が要求されるシステム等においては、リード処理を優先するように設定されるべきである。このことから、前記負荷分散制御部12におけるデータ及び冗長データの配置アルゴリズムは、前述のアクセス履歴管理情報及びユーザの指示情報から、再設定可能であることが望ましい。

【0092】(第3実施例) 図11は、本発明の第3実施例に係るディスクアレイ装置におけるデータの配置の1例を示す図である。本実施例のディスクアレイ装置のブロック図は、図1または図8のブロック図と略同様のものであり、図示していない。本実施例では、前記第1、第2実施例と異なり、前記ディスクアレイ装置2は冗長データを持たず、構成する個々のディスク装置仕様

に基づき、データの配置の最適化を行うことによって、システム性能の向上を実現するようになっている。

【0093】図11は、ディスク装置仕様の異なる3種類のディスク装置(43, 44-1, 44-2, 45-1, 45-2)からディスク装置群4を構成する場合における、データ配置の1例を示したものである。

【0094】図12は、図2と同様に、本実施例の前記構成ディスク仕様管理部13における、個々の構成ディスク仕様管理情報の1例を示す図である。図12は図11に対応しており、図11のディスク装置43がI D=AAA'に、ディスク装置44-1がI D=BBB'に、ディスク装置44-2がI D=CCC'に、ディスク装置45-1がI D=DDD'に、ディスク装置45-2がI D=EEE'にそれぞれ対応している。

【0095】本実施例において、ディスク装置群4を構成する個々のディスク装置仕様を、以下のように仮定した場合を例にとって、データの配置について説明する。

【0096】[ディスク装置仕様] I/F I D=0, CH I D=4のディスク装置45-2(I D=EEE')を基準ディスク装置仕様とする場合、I/F I D=0, CH I D=3のディスク装置45-1(I D=DDD')は、基準ディスク装置と同一仕様、I/F I D=0, CH I D=2のディスク装置44-2(I D=CCC')は、基準ディスク装置の2倍の仕様、I/F I D=0, CH I D=1のディスク装置44-1(I D=BBB')は、基準ディスク装置の2倍の仕様、I/F I D=0, CH I D=0のディスク装置43(I D=AAA')は、基準ディスク装置の4倍の仕様とする。

【0097】[データ配置] 本実施例においても、前述の第1、第2実施例と同様に、ストライプ数の設定によって負荷分散制御を行うものである。3種類のディスク装置仕様が1:2:4の比で示されることから、ストライプ数も同様に1:2:4を設定し、設定ストライプ数に対応し連続するデータストライプ(例えば、ストライプ#0~#3)は同一ディスク装置に格納するようにしている。

【0098】ランダムアクセスにおいて、このような個々のディスク装置の総データストライプ数比と性能比とを同等とするディスクアレイ装置2を構築することによって、ディスク装置43へはディスク装置45-1, 45-2に対し4倍、ディスク装置44-1, 44-2へはディスク装置45-1, 45-2に対し2倍のアクセス発生が期待される。すなわち、全てのディスク装置に均等にデータストライプを配置する従来方式では、全てのディスク装置に対して等しい確率でアクセス要求が行われる場合、ディスク装置43, 44-1, 44-2の性能が活かしきれず任意の処理時間内に5件のI/O要求しか処理されないのに対し、本実施例によれば、上記任意の処理時間内に10件のI/O要求が処理可能とな

り、2倍のシステム性能の向上が図られる。

【0099】図13は、本実施例におけるランダム/シーケンシャルリードアクセス時の性能比較を示す1例であり、図中の波線はシーク/回転待ち時間を、実線はデータ転送時間をそれぞれ示している。図13において左側が、ランダムリードアクセスを行なった場合の性能を示し、図13において右側が、シーケンシャルリードアクセスを行なった場合の性能を示している。

【0100】シーケンシャルアクセスにおいて、個々のディスク装置の総データストライプ数比と性能比とを同等とするディスクアレイ装置2を構築し、連続するデータストライプを同一のディスク装置に格納することによって、データ転送性能を向上することが可能である。すなわち、全てのディスク装置に均等にデータストライプを配置する従来方式では、ディスク装置43, 44-1, 44-2の性能が活かしきれず、ディスク装置45-1, 45-2の5倍のデータ転送性能しか実現できないのに対し、本実施例によれば、ディスク装置45-1, 45-2の10倍のデータ転送性能が実現可能となり、2倍のシステム性能の向上が図られる。

【0101】以上の通り、冗長データを持たない本実施例においても、前記ディスクアレイ装置2を構成する個々のディスク装置仕様に基づき、データの配置の最適化を行うことによって、システム性能の向上を実現することが可能となる。

【0102】(第4実施例) 図14は、本発明の第4実施例に係るディスクアレイ装置のブロック図である。本実施例でも、ディスクアレイ装置2は、アレイ制御部10と、ディスク装置群4とから構成され、アレイ制御部10のメイン制御部は、前記第2実施例がもつ機能を少なくとも具備したものとなっている。また、ディスク装置群4は、複数のディスク装置42から構成される基本ユニット48、及び複数のディスク装置41から構成される増設ユニット49から構成されている。なお、図14において、18は複数の増設ディスクインタフェース制御部である。

【0103】上記したディスクアレイ装置2の構成において、前記基本/増設の両ユニット48, 49を構成するディスク装置仕様が異なる場合、ディスク装置(41, 42)仕様に基づき、データ、冗長データの配置の最適化を行うことによって、ディスク装置増設時においても、システム性能の向上を実現するようにしたものである。

【0104】本実施例のデータ及び該データに対応する冗長データの配置は図示していないが、例えば、前記図6におけるディスク装置41-0, 41-1を増設ユニット49のディスク装置41と見做し、図6におけるディスク装置42-2, 42-3, 42-4を基本ユニット48のディスク装置42と見做せば、本実施例のデータ及び該データに対応する冗長データの配置の1例とな

る(但し、図6とはディスク装置の数が異なる)。この場合、基本ユニット48 或いは増設ユニット49を構成するディスク装置仕様は、それぞれ同一のものとしていることになる。

【0105】本実施例のアレイ制御部10のメイン制御部11は、現在未接続の増設ディスクインタフェース制御部18の何れかに、新たにディスク装置41が接続・増設されると、このディスク装置の増設及び増設されたディスク装置仕様を検出し、この増設ディスク装置仕様情報を、前記した構成ディスク仕様管理部13に新規に追加格納してこれを管理する。

【0106】すなわち、新たにディスク装置41が接続・増設されると、アレイ制御部10のメイン制御部11は、①. 新規の増設ディスク装置の検出、②. 増設ディスク装置仕様の獲得、③. 負荷分散制御方式の決定、④. データの格納処理(所定の2つのRAIDの構築、データストライプの再配置) を実行する。

【0107】そして本実施例においても、前記第2実施例と同様に、アクセス履歴管理情報及び構成ディスク仕様管理情報に基づき、アクセス頻度の高いデータ或いは該データに対応する冗長データを、より高性能なディスク装置から構成されるユニットに格納する再配置を実行することによって、システム性能の向上を期待することが可能となる。

【0108】以上、本発明を図示した実施例によって説明したが、当業者には本発明の精神を逸脱しない範囲で種々の変形が可能であることは言うまでもなく、例えば、メイン制御部11が上位装置(ホスト1)において実現され、上位インタフェース制御部16は上位装置の内部バスに接続され、更に、バッファ部15が上位装置のメモリ部を用いて構成されるものであっても、差し支えない。

【0109】

【発明の効果】以上のように本発明によれば、仕様の異なる複数のディスク装置から構成されるディスクアレイ装置において、個々のディスク装置仕様に基づきデータ、冗長データの配置の最適化を行うことによって、ディスクアレイ装置を構成する個々のディスク装置仕様に対応したシステム性能を実現することが可能となる。

【0110】すなわち、ディスクアレイ装置を構成するディスク装置が障害等の発生によって、高性能なディスク装置に交換された場合、この交換された高性能ディスク装置の性能を活用可能とすることによって、システム性能の向上が可能となる。

【0111】また、レベルの異なる複数のRAIDレベルから前記ディスクアレイ装置を構築可能とすることによって、従来方式では使用されない領域の有効活用が可能であり、この領域をデータ或いは冗長データの格納に使用することによって、従来と同等以上の冗長度を確保した高信頼な記憶装置の構築が可能となる。

【0112】また、前記ディスクアレイ装置を構成する個々のディスク装置仕様に応じて、データストライプ数比を任意に設定し、前記高性能なディスク装置に対するアクセス発生頻度の期待値を高めることによって、システム性能の向上を図ることが可能である。

【0113】また、前記ディスクアレイ装置を構成する個々のディスク装置に対する過去のアクセス履歴情報に基づき、データストライプ或いは冗長ストライプの再配置を実行可能とすることによって、前記高性能なディスク装置に対するアクセス発生頻度の期待値を高めることが可能となり、システム性能を更に向上させることが可能である。

【図面の簡単な説明】

【図1】本発明の第1実施例に係るディスクアレイ装置のブロック図である。

【図2】本発明の第1実施例によるディスクアレイ装置で用いられる構成ディスク仕様管理情報の1例を示す説明図である。

【図3】本発明の第1実施例によるディスクアレイ装置における、データ、冗長データの配置の第1例を示す説明図である。

【図4】本発明の第1実施例によるディスクアレイ装置と従来のディスクアレイ装置との性能比較の1例を示す説明図である。

【図5】本発明の第1実施例によるディスクアレイ装置における、データ、冗長データの配置の第2例を示す説明図である。

【図6】本発明の第1実施例によるディスクアレイ装置における、データ、冗長データの配置の第3例を示す説明図である。

【図7】本発明の第1実施例によるディスクアレイ装置における、データ、冗長データの配置の第4例を示す説明図である。

【図8】本発明の第2実施例に係るディスクアレイ装置のブロック図である。

【図9】本発明の第2実施例によるディスクアレイ装置で用いられるアクセス履歴管理情報の1例を示す説明図である。

【図10】本発明の第2実施例によるディスクアレイ装置における、データ、冗長データの配置の1例を示す説明図である。

【図11】本発明の第3実施例によるディスクアレイ装置におけるデータの配置の1例を示す説明図である。

【図12】本発明の第3実施例によるディスクアレイ装置で用いられる構成ディスク仕様管理情報の1例を示す説明図である。

【図13】本発明の第3実施例によるディスクアレイ装置における、ランダムリードアクセス時とシーケンシャルリードアクセス時との性能比較の1例を示す説明図である。

24

17 ディスクインタフェース制御部

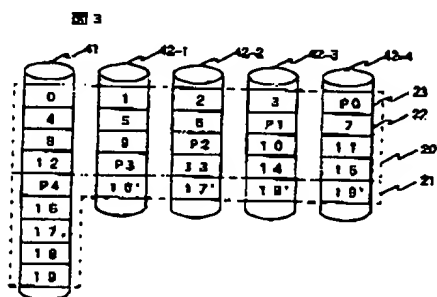
18 増設ディスクインタフェース制御部

- ```

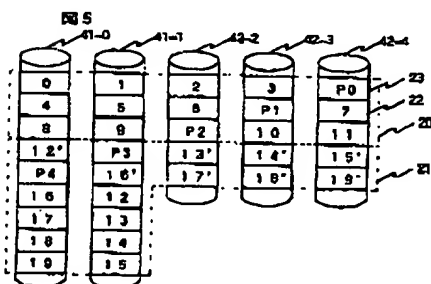
20 RAI D5 領域
21 擬似RAI D1 領域
22 データストライプ
23 パリティデータストライプ
41, 41-0, 41-1, 41-2, 42, 42-
1, 42-2, 42-3, 42-4, 43, 44, 44
-1, 44-2, 45, 45-1, 45-2 ディスク
装置
48 基本ユニット
49 増設ユニット

```

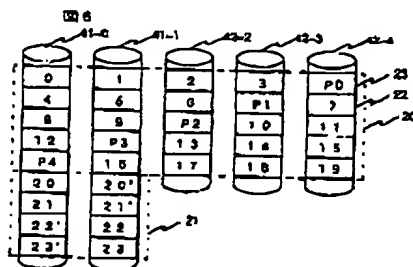
【 図3 】



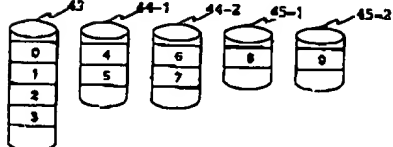
【 図5 】



【凶6】



**DE 2 7**



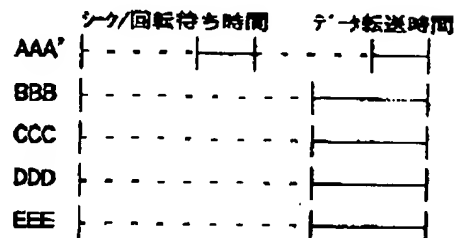
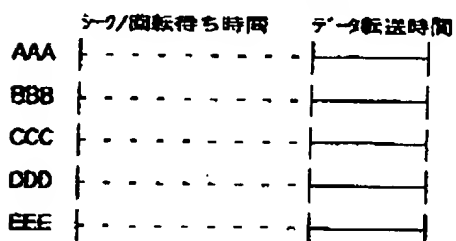
【 図2 】

図 2

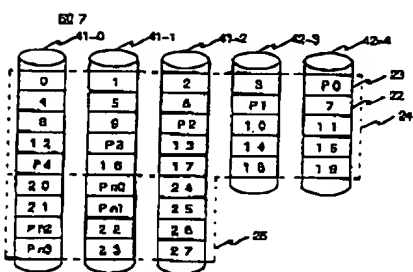
| 31<br>SCSI (I/F) ID | 32<br>Channel ID | 33<br>Disk ID | 34<br>Capacity | 35<br>Seek Time | 36<br>RPM | 37<br>Transfer |
|---------------------|------------------|---------------|----------------|-----------------|-----------|----------------|
| 0                   | 0                | AAA'          | xx1            | ss1             | rr1       | tt1            |
| 0                   | 1                | BBB           | xx0            | ss0             | rr0       | tt0            |
| 0                   | 2                | CCC           | xx0            | ss0             | rr0       | tt0            |
| 0                   | 3                | DDD           | xx0            | ss0             | rr0       | tt0            |
| 0                   | 4                | EEE           | xx0            | ss0             | rr0       | tt0            |
| ?                   | ?                | ?             | ?              | ?               | ?         | ?              |

【 図4 】

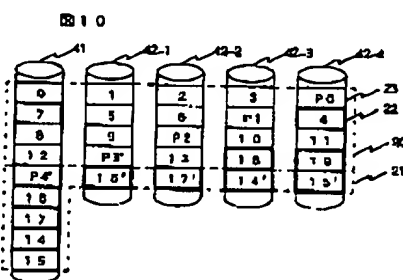
図 4



【 図7 】



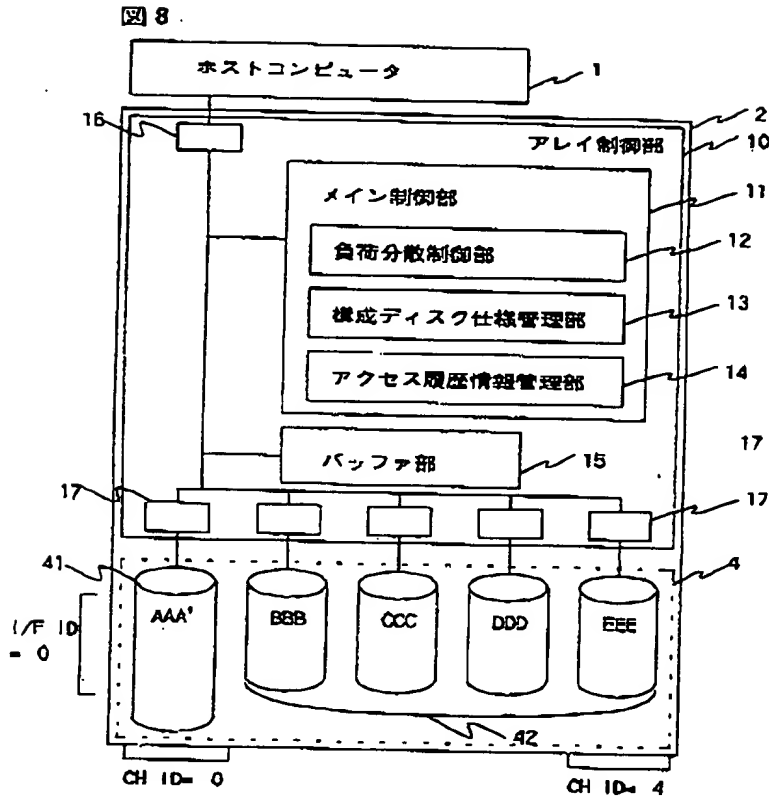
【 図10 】



【 図12 】

| 31<br>SCSI (I/F) ID | 32<br>Channel ID | 33<br>Disk ID | 34<br>Capacity | 35<br>Seek Time | 36<br>RPM | 37<br>Transfer |
|---------------------|------------------|---------------|----------------|-----------------|-----------|----------------|
| 0                   | 0                | AAA'          | xx01           | ss01            | rr01      | tt01           |
| 0                   | 1                | BBB'          | xx02           | ss02            | rr02      | tt02           |
| 0                   | 2                | CCC'          | xx02           | ss02            | rr02      | tt02           |
| 0                   | 3                | DDD'          | xx03           | ss03            | rr03      | tt03           |
| 0                   | 4                | EEE'          | xx03           | ss03            | rr03      | tt03           |

【 図8 】



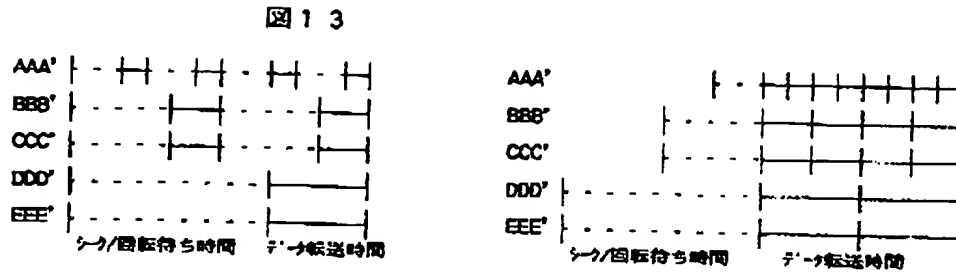
【 図9 】

図 9

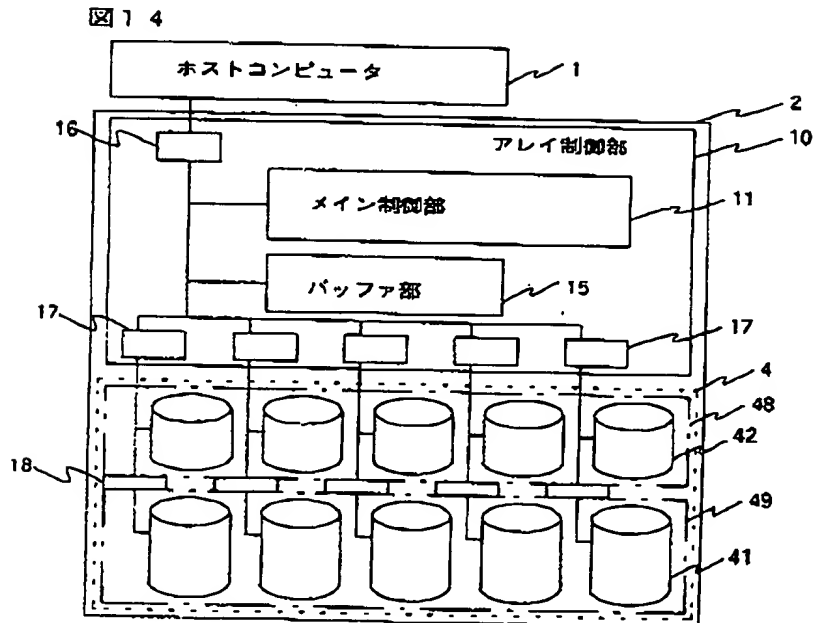
| Strip # | アクセス傾向   |          |          | Strip # | アクセス傾向   |          |          |
|---------|----------|----------|----------|---------|----------|----------|----------|
|         | READ     | WRITE    | Total    |         | READ     | WRITE    | Total    |
| 0       | RD0 ( )  | WR0 ( )  | TT0 ( )  | 15      | RD15 (4) | WR14 (2) | TT14 (2) |
| 2       |          |          |          | 16      | RD16 ( ) | WR16 (3) | TT16 ( ) |
| 4       | RD4 (1)  | WR4 ( )  | TT4 ( )  | 17      | RD17 ( ) | WR17 (4) | TT17 ( ) |
| 2       |          |          |          | 2       |          |          |          |
| 14      | RD14 (3) | WR14 (1) | TT14 (1) | 19      | RD23 ( ) | WR23 ( ) | TT23 ( ) |
|         |          |          |          | TOTAL   | RD1      | WR1      | TOTAL    |

( )内は、アクセス発生頻度順位

【 図1 3 】



【 図1 4 】



フロント ページの続き

(72) 発明者 松本 純  
 神奈川県川崎市麻生区王禅寺1099番地 株  
 式会社日立製作所システム開発研究所内

(72) 発明者 岩崎 秀彦  
 神奈川県小田原市国府津2880番地 株式会  
 社日立製作所ストレージシステム事業部内

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☒ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**